DOCUMENT RESUME

ED 106 405 UD 015 141

AUTHOR Cain, Glen G.
TITLE A Review of "Socioeconomic Background and Achievement" by Otis Dudley Duncan, David L. Featherman, and Beverly Duncan. Discussion Papers No. 188-73.
INSTITUTION Wisconsin Univ., Madison. Inst. for Research on Poverty.
SPONS AGENCY Office of Economic Opportunity, Washington, D.C.
REPORT NO IRP-DP-188-73
PUB DATE Jan 74
NOTE 27p.

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS Attitudes; Critical Path Method; Demography; Family Background; Income; Intelligence Differences; Motivation; *Occupational Mobility; Peer Groups; Racial Differences; *Research Reviews (Publications); School Role; *Statistical Analysis; Working Women

ABSTRACT

The book under review deals with models which identify and measure factors determining a man's educational, income, and occupational attainments. The factors include parental background, number of siblings, national origin, race, intelligence, aspirations, motivations, peer group influences, schools, the influence of wife's characteristics, first job, age at first job, migration, fertility, and child spacing. These models and their estimation constitute a body of research that should be interesting and important to a variety of social scientists and highly relevant to the questions of the causes and consequences of poverty. Although the mathematical and statistical techniques used are sometimes complex, even the non-technical reader can learn much from the methodological discussions. Criticisms are presented under four headings: (a) a vagueness of the specific purposes of their analyses; (b) a selection of a mode of presenting their findings which hinders translating them into an understandable language, and thereby impedes useful and practical applications of their results; (c) excessive complexity of the models and sometimes confusing exposition; and (d) a number of dubious rationalizations of the structures of their models and a number of dubious interpretations of some of their empirical results. (Author/JM)

188-73

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

A REVIEW OF SOCIOECONOMIC BACKGROUND AND ACHIEVEMENT

by Otis Dudley Duncan, David L. Featherman, and Beverly Duncan

Reviewed by Glen G. Cain

UNIVERSITY OF WISCONSIN - MADISON

ED106405

A REVIEW OF SOCIOECONOMIC BACKGROUND AND ACHIEVEMENT

by

Otis Dudley Duncan, David L. Featherman, and Beverly Duncan

Reviewed by Glen G. Cain

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

January 1974

UD 015141

This work was supported in part by funds granted to the Institute for Research on Poverty, University of Wisconsin, pursuant to the provisions of the Economic Opportunity Act of 1964. The author is grateful to David L. Featherman for helpful comments on an earlier draft and to Arthur S. Goldberger for helpful discussions at several stages, but neither colleague necessarily agrees with any part of the current version.

ABSTRACT

The book under review deals with models which identify and measure factors determining a man's educational, income, and occupational attainments. The factors include parental background, number of siblings, national origin, race, intelligence, aspirations, motivations, peer group influences, schools, the influence of one's wife's characteristics, first job, age at first job, migration, fertility, and child-spacing. These models and their estimation constitute a body of research that should be interesting and important to a variety of social scientists and highly relevant to the questions of the causes and consequences of poverty. Although the mathematical and statistical techniques used are sometimes complex, even the nontechnical reader can learn much from the methodological discussions. The review adopts a critical posture in the hope of suggesting ways in which this type of research could reach a wider audience and contribute more effectively to the application of social science to coping with social problems.

A REVIEW OF SOCIOECONOMIC BACKGROUND AND ACHIEVEMENT

by

Otis Dudley Duncan, David L. Featherman, and Beverly Duncan

Reviewed by Glen G. Cain

This book is impressive, almost overpowering, in the scope of its subject matter, methodological erudition, technical virtuosity, and substantive empirical findings. In general terms, the book deals with the construction of models which identify factors determining a man's income and educational and occupational attainments. Clearly, these outcomes are largely responsible for one's social achievements, status, even well-being; and for this reason the models should be of intense interest to scholars among several disciplines in the social sciences. The following list of variables which are examined as determinants of these achievement outcomes should further whet the appetites across disciplines: parental background, composition of family of origin, national origin, race, intelligence, aspirations, motivations, peer group influences, schools, the influence of one's wife, first job, age at first job, migration, fertility, and child-spacing--some of which variables are, of course, interesting outcomes in their own right.

The models developed consist of systems of equations. The outcome variables are dependent variables which are systematically determined by three types of independent variables: background variables, such as race and father's education; intervening variables, such as intelligence and peer group influences; and career contingencies, such as age at first job and migration. The system of equations represents a temporal sequence of outcomes, which means that a "first" outcome variable, like educational attainment, can also be an independent variable in the second equation in

which occupational attainment is the outcome. This recursive structure allows estimating the parameters or relationships among the variables by ordinary, least-squares regression methods.

The basic data source is a large survey of males aged 25-64 in 1962. The variables are scaled in standard units, i.e., each variable is subtracted from its sample mean and divided by its sample deviation. The parameter estimates are called path coefficients (or beta-weights).

In the first three chapters, the authors present the form of the model, explain the statistical estimation methods, and analyze a simple, "basic" model that had previously been the main focus of this book's predecessor (Blau and Duncan 1967). The succeeding six chapters expand upon the basic model with variations in independent variables selected, different data sources, and more elaborate models (discussed below).

The mathematical and statistical techniques displayed are considerable and, unfortunately, are likely to repel many readers who could profit from the book. I will argue below that some of these barriers to a wider readership could have been avoided; initially, however, it should be emphasized that even the non-technical reader can learn much from the methodological discussions, the empirical findings, and the accompanying interpretative discussions.

The authors' style of presentation is a mixture of modesty and boldness. They are explicit in inviting the reader to question their work and to offer alternatives. I have accepted this invitation and will criticize, hopefully constructively, more than I will praise or even summarize.

I present the criticisms under the headings of what I perceive to be four related shortcomings of the book: (a) a vagueness of the specific purposes of their analyses; (b) a selection of a mode of presenting their

findings which hinders translating them into an understandable language, and thereby impedes useful and practical applications of their results; (c) excessive complexity of the models and sometimes confusing exposition; (d) a number of dubious rationalizations of the structures of their models and a number of dubious interpretations of some of their empirical results.

1. PURPOSES

Because the models and estimating techniques are often complex, a clear and explicit statement of what specific purposes and uses the empirical results might serve would help to motivate and guide the reader. Specifically, consider the purpose of informing community or governmental decision-makers about policy actions. Assume that the process or "experiment" under consideration is upgrading the occupational attainment and/or affecting the family size of a cohort of young married people, wherein one outcome of interest is the effect of these changes on the educational attainment of the children. In this case it would be necessary to ask two preliminary questions about the models in this book that include these variables: (a) Is the process which generated the reported empirical results sufficiently similar to the contemplated policy action; and, (b) if so, what are the measured effects of the variables on the outcomes in terms that the policymaker can understand?

By comparison with policy formulation, the purposes of prediction in a context without deliberate intervention is somewhat less demanding. We might, for example, be interested in whether the observed variation in father's occupational attainment and family size yield accurate predictions of the educational outcomes of the children. (Let us set aside the important question of what we mean by "accurate," since this would require that we specify what the predictions are to be used for, what the costs of

making mistakes are, and so on.) Speaking loosely, we may say that accurate predictions may be obtained if the process of the social and economic system which generated the existing data will continue to operate in the same way in the future. More specifically, the question is: How stable is the covariance structure of the relevant set of variables over time?

The relevance of the stability of the covariance structure may be illustrated with the basic model demonstrated in Chapter 3 in which father's occupation, father's education, and the number of siblings are determinants of the son's educational attainment. We can assume further that the wealth--or, in economists' terms, the permanent or normal income--of the parents is also a determinant of the educational attainment of the children. (Current income as measured for a single year is not designated because it is often a weak proxy for wealth or normal income.) The income (or wealth) variable is omitted from the authors' model. Does this lead to biased estimates of how educational attainment will respond to changes in the three background variables over time? The answer is "no" if the covariance structure among the three variables and income remains the same. It is reasonable to believe that the effect of wealth previously was partially "captured" in its relations--the "old" covariance structure--to the three background variables. The variation in wealth that was independent of these three background variables was "captured" in the error term of the model. But the ability of the three background variables to yield unbiased predictions depends on that covariance structure remaining the same in the new context to which it is applied.

Although the authors discuss the general issue of omitted variables, the issue cannot be evaluated or even become interesting, unless the purposes of the model are made explicit. What is the process or experiment

about which inferences are to be made? If the purpose were to know the effect of a policy action, such as changing ultimate family size (perhaps by the introduction of a family planning program), then we see explicitly that the intervention will introduce variation in one background variable and not in the other three (i.e., father's occupation, education, and wealth). Here the demands for unbiased measurement of the effect of a change in family size are very strict, because we must know its "net" or "partial" effect, holding constant all three variables, including wealth.

Clearly, many other variables besides wealth could become relevant omitted variables, but, to repeat, the purpose of the models and their domain of application must be known before evaluating the question. The authors do not address this question explicitly, although my impression is that historical prediction is the purpose for which their models are best suited.

2. TRANSLATING, INTERPRETING, AND APPLYING THE EMPIRICAL RESULTS

Path Coefficients

The use of path coefficients limits the usability of the empirical results for many purposes. To illustrate this, we can assume that the reader wants to know the estimated quantitative relation between years of schooling of the father and the occupational achievement of the son. The reader must first transform a standard-deviation unit of the two variables into their raw units--i.e., actual years of schooling and actual occupational scores. If the regression coefficients of the raw units were presented, it would not be necessary to make special calculations to determine that, for example, an extra year of the father's education (say, having 9 or 13 years of schooling completed instead of 8 or 12), implies

the prediction that the son will advance in occupational attainment by two points. (A linear and additive specification of the independent variables is the functional form for the models throughout the book.)[1]

After applying the path coefficients to the transformed units, the user will still have to translate the change in occupational scores--two points in the example being discussed--into another unit which permits an understanding and communication of the result. One can determine that a two point gain means advancing from a brakeman to a switchman, a bartender to a waterworks man, or miner to shoemaker--to take three examples of pairs of occupations separated by two points.[2]

The user will need, however, to persevere further. The two-point change reflected only the direct effect of father's education on son's occupation. In addition one must calculate the indirect effect--via the direct effect of the father's education on the son's education, and then of the direct effect of the son's education on the son's occupational achievement. Transformations of the path coefficients into raw units and then of the raw units of occupational scores into "representative" changes in occupations are again required.

One alternative to using the occupational categories as a way of sensing a meaning of the results is to convert the occupational scores into units of annual earnings. This conversion could be accomplished in several ways, but this device is suggested only as a complement and not as a substitute indicator of occupational change. There is a need for some indicator that would be commonly understood. Indeed, if the authors would provide such translations, they would no doubt feel compelled to comment on the practical significance of the quantitative relations. This would be beneficial, since they are eminently qualified to do so. As it now

stands, much of the significance of the book consists of piecemeal empirical findings, which are difficult to translate into predictions or evaluations of the social processes, and of findings which are mostly of methodological interest.

Measuring the Dependent Variable

It is unfortunate that the authors provide no resumé of the operational definition of the occupational score variable or of its strengths and weaknesses. The reader can pursue the cited literature on occupational measurements, but ploughing through this is neither convenient nor adequate. On page seven the authors simply state that the respondent's occupation is "the ultimate outcome of the whole process" and that, "the letter Y stands for the variable, occupational socioeconomic status, as measured on the scale developed by Duncan." The reference (Duncan 1961) reveals that a predicting equation was constructed in which an occupational prestige score (Y') was regressed on age-adjusted measures of income (I) and educational attainment (E) in the occupation. (I is the percent in the occupation which earned $3,500 or more in 1949, and E is the percent in the occupation with 12 or more years of schooling completed, with both variables scaled in standard units.) Forty-five occupations with prestige scores based on a 1947 survey were used, and the 1950 census provided the income and education variables. The equation is:

$$Y = \hat{Y}' = a_0 + a_1 I + a_2 E.$$

The estimated relationship in the Duncan article cited is:

$$Y = -6.0 + .59I + .55E$$

Therefore, the SES value, Y, is determined for any occupation for which the requisite information about I and E exists.

We see that Y could be defined, with probably little change in results, as the simple sum of the percent earning more than $3,500 and the percent with 12 or more years of schooling in the occupation, since .59I + .55E is approximately equal to .57(I + E), and the constant .57, as well as the constant -6.0, is irrelevant to the performance of Y in the regression models.

It should also be pointed out that because Y is defined in standard units, a one unit change in Y will mean different amounts of changes in the SES scores whenever the variance of SES differs from one sample to another. Thus, the user should be aware that a dependent variable, for example, son's educational attainment, may be more responsive to, say, SES of the father in Study A than in Study B even though the coefficient of father's Y (equal to SES in standard units) is smaller in Study A than in Study B. Clearly, the issue is interesting only if the units of the occupational scores have some meaning to readers--if only in the limited way that IQ scores or temperatures have meaning. If there is no cardinal meaning to the units, then, of course, any statistic which measures the sign of the relationship--regression coefficient, correlation coefficient, or beta-weight--would serve equally well. I assume that there is some interest in the actual SES scores, and I would insist there is interest in the "natural" units which measure schooling, income, age, and other variables used in the book.

Several other questions may be raised about the key variable, Y.

(1) Would the equation for Y, which was based on the 1947 and 1950 data, change substantially if it were re-estimated with more recent and

more abundant occupational prestige scores and the 1960 census data? Are there advantages in using the more recent measure in a study of occupational attainment in 1962? What differences would this have made? What criteria could determine whether the differences are important or unimportant?

(2) A conceptual statistical problem with the estimating equation is the apparent inconsistency of the estimate a_1, the coefficient of I. If the economic theories of a trade-off between income and prestige are correct (as the authors themselves imply on p. 243), then changes in prestige will, other things equal, change income. The mechanism is that an increase in prestige will attract new entrants to the occupation, and the augmented supply will depress wages and earnings. The reverse process would occur if the prestige of the occupation underwent an autonomous decline. A dependence between the error term of the equation and the income variable results, and this will produce an inconsistent estimator of a_1.

(3) A comparison of the SES score and the actual prestige score on pages 48-49 creates in my mind more uncertainty about the properties of the score. The occupations listed for fathers and sons in the Detroit Area Survey are assigned the prestige score, Y', and the SES score, Y.[3] The same scores are, of course, assigned to any given occupation irrespective of whether it is the son's or father's occupation. Thus, the regression between Y and Y', e.g., $Y' = a + bY + u$, will yield approximately the same intercept and slope coefficients, whether fathers or sons are the source for the sample of occupations. In particular, the error term, u, in the regression will tend to be positive or negative for any given occupation, irrespective of whether the error term is computed from the

sons' or fathers' regression. Given a correlation between father's and son's occupations, it, therefore, seems untenable to assume uncorrelated errors as between Y and Y' for fathers and sons. Yet the authors make this unwarranted assumption to justify a series of calculations examining the issue of whether Y or Y' is the "truer" measure of occupational status and prestige.

(4) The definition of the SES score suggests that a "mechanical" correlation exists between it and the educational attainment of the respondent; and the correlation will be higher the more the respondent's occupation is educationally homogeneous. Mitigating this correlation is the fact that the educational component of the SES score is the percent with 12 or more years of schooling completed, whereas the respondent's education is measured in years of schooling completed.

(5) Finally, the operational definition of the SES variable may explain why the relation between father's education and son's education is less when the SES score is used than when the actual prestige score is used. The reason is that the father's SES score has a built-in educational component, so the net effect of the father's education on the son's education is diluted when the father's SES score is included--given the positive correlation between a person's education and the educational measure of his occupation.

These remarks are not made contentiously. Indeed, they are motivated by the technical detail about scales of measurement which the authors have scrupulously provided--in a scant five pages! The remarks are intended to re-open the issues of the meaning and interpretation of occupational scores.

3. EXCESSIVE COMPLEXITY OF THE MODELS

Chapter 6 involves an investigation of aspirations and motivations as intervening variables. It is the longest, the most difficult, and to me the most confusing chapter in the book. Complexities arise when latent (or unobserved) variables and simultaneously-determined (or "feedback") variables are introduced. Although these formulations are sometimes essential for a correct modeling of the process under investigation, I believe they produce in this chapter a number of redundant and uninterpretable empirical results, and they serve to illustrate my general criticism of "excessive" complexity.

(1) An early model in the chapter provides an example of how a much simpler presentation can achieve the same conclusions as a more complex model. Consider Model I:

$$x_3 = p_{31}x_1 + p_{32}x_2$$

$$x_6 = p_{63}x_3 + p_{61}x_1 + p_{62}x_2$$

where x_3 = educational plans of the student; x_1 = father's SES; x_2 = student's I.Q.; x_6 = actual educational attainment; the p's are path coefficients. The error terms are omitted from the equations.

The authors offer a new model which, I claim, is in effect Model II:

$$x_3 = p_{31}x_1 + p_{32}x_2$$

$$x_6 = q_{63}x_3 + q_{62}x_2$$

which incorporates the authors' assumption that $p_{61} = 0$. The authors

conclude, by a route discussed below, that $q_{63} > p_{63}$. But isn't it all but self-evident that the effect of the educational plans of the student on his educational outcomes will be larger in an equation in which father's SES (x_1) has been omitted?[4] They claim, furthermore, that their alternative model "requires a 'sleeper effect' of intelligence to account for the correlation of attainment with background." Again, the required sleeper effect is self-evident from the imposed assumption that background (x_1) has a zero direct effect.

The obviousness of these (and some other) conclusions are, however, obscured by a confusing substitution of symbols; namely, $x_a = p_{3a}x_3 + e$, where x_a is called the latent decision, and e may be considered an error term that is uncorrelated with other variables. The authors' model can, in fact, be derived by simply substituting in Model II the three unknown quantities, $(x_a-e)/p_{3a}$, for the observed variable, x_3. A number of manipulations and calculations will then only produce a display of p_{ai} and p_{ja} coefficients instead of the p_{3i} and q_{j3} coefficients available in Model II. While there may be methodological interest in an estimate of a reliability coefficient for p_{3a}, the limited practical information available from Models I and II is obscured by the presentation of results in terms of a "latent" variable.

(2) Single-equation and recursive models are not strictly appropriate when there is mutual determination, or simultaneity, among variables in a model. As the authors point out, contemporaneous measures of "motivation" and "occupational achievement" are likely to reflect mutual causation. For me this fact undercuts any substantive interest in the authors' ingenious, but involved, re-analysis (pp. 116-130) of a model by Harry J. Crockett, which ignores this feedback relation. However, readers who are

willing to accept the assumptions that the measure of motivation (Thematic Apperception Tests) is both a predetermined and accurate measure of the motivation that applied when decisions were made years earlier, will find the authors' re-analyses of more than methodological interest.

In a subsequent model in this chapter the authors confront a similar problem in examining the relationships among five motivational variables: (a) "subjective achievement" (SA), (b) "importance of getting ahead" (IGA), (c) a latent variable, "ambition" (A), which is a substitute for IGA, (d) "commitment to work" (CW), (e) a latent variable, "work orientation" (WO), which is substituted for CW; and, finally, (f) a series of conventional variables: father's occupation, siblings, education, occupation at marriage, income at marriage, current occupation, and current income. I confess that I could not fully follow the development and estimation of some of the models with these variables, so the following comments are tentative.

Two recursive models are offered. The first is called a "naive" recursive model in which father's occupation, siblings, and IGA determine education; all four in turn determine CW; and all five in turn determine the remaining conventional variables. In a second model the unobservables, A and WO, replace the observables, IGA and CW. The introduction of A and WO make the model underidentified, but empirical estimation goes forward by assuming various values for path coefficients linking A to IGA and WO to CW.[5]

After an extraordinarily elaborate (even for this book) set of calculations, the latent-variable model yields empirical results, many of which are similar to the naive model, others change signs as different assumed reliabilities are used, and others are simply anomalous. (For example, both A and IGA are negatively related to education; the only significant

coefficient of father's occupation on son's occupation is negative; and so on.)

To be fair, the authors label this section exploratory, and they admit to a certain arbitrariness in their causal-recursive ordering of IGA, CW, and SA. (Let me note that a component of CW and, indirectly, A, is "the work I do is one of the most satisfying parts of my life," and yet CW and A are specified as causally prior to "current occupation" and "current income"!) The most important questions I would raise about this section, however, do not concern the empirical results; rather they are related to the question of purpose, mentioned above. The authors state their central concern to be testing whether motivations "are 'relevant and important' to social mobility" (p. 134). What is the context in which this issue is posed? "Important" by what criterion and in the context of what process?

(3) Educational and occupational aspirations are also examined in Chapter 6 as intervening variables in a system in which father's SES and IQ are background variables. To these variables the authors add a latent motivation variable, M. I examined in detail the first (WISC-I) of two models, and I will merely assert here that I believe this addition of the latent variable again yields no information that could not directly, more easily, and more understandably be derived from the observed variables and the same restrictions on the model that the authors must make to convert the underidentified system (with M) into an estimatable one. (I will be glad to send any interested reader my documentation for this assertion.)

Incidentally, the necessary restrictions are not altogether palatable, because they require that, (a) the respondent's educational aspirations depend on his occupational aspirations and on his father's SES but not on

his own I.Q.; and (b) that his occupational aspirations depend on his father's SES and his I.Q. but not on his educational aspirations.[6] Regarding (a) it seems unreasonable to assume that a person's I.Q., which partially determines his success in school, won't directly affect his aspirations for further schooling. With respect to (b) it also seems unreasonable to assume that occupational aspirations are not influences by educational aspirations. There are probably many young people who do well in school, who aspire to college, and who will, as a consequence of their educational aspirations, aspire to an occupation that befits a college graduate. Thus, the system of equations determining the two aspiration variables appears hopelessly underidentified. Perhaps it is a matter of taste, but I believe it is more instructive to readers simply to point this out, rather than to expend so much effort--some pedagogical rewards notwithstanding--in attempting to square the circle.

4. INTERPRETATIONS OF THE MODELS

The variety of models and abundance of empirical findings expressing quantitative relationships among variables guarantee a rich learning experience for the attentive reader. No doubt many readers will engage in silent disputations with the authors about their models and the interpretations made of the empirical results. I will confine my long list of potential comments and criticisms to a few cases which illustrate some general points.

One source of challenge to some of the authors' interpretations stems from the economist's view of the household as a decision-making unit. This view leads to a distinction between endogenous variables, which represent the outcomes of the decisions, and exogenous or predetermined variables, which impinge on these decisions. These categories of variables do not necessarily conflict with the categories the authors use, but the

framework may point to some different interpretations. For example, in contrast to the authors' model in which educational attainment, age of first job, status of first job, and, in places, age at marriage are ordered in a particular temporal sequence, the economist's view might be that these contingency variables comprise a common set of mutually interdependent choices. Thus, the decision to pursue a college career is simultaneously made along with the decisions to enter a first job at around age 22, to enter, say, the engineering occupation, and to marry at around age 23 or 24. The term "simultaneous" should be thought of in this context as referring to an inextricable set of decisions which have been meshed over a period of time prior to the outcomes being analyzed. Viewed in this perspective, the authors' temporal ordering, in which some of these variables appear as exogenous right-hand side variables presumed to be uncorrelated with the error term in the model, is not correct. The ordinary least-squares regression estimates of the recursive system would produce inconsistent and biased estimates of the parameters in the simultaneous equations system, and if the latter system is the correct modeling of the process. then only these estimates are relevant.[7] I should add that in my opinion economists have not been very productive in empirically estimating these processes, so that there is little evidence to demonstrate which model--recursive or simultaneous--is "better." Recursive models do have the virtue of being less complicated.

Economists have more successfully employed the household decision-making model to analyze behavior of wives, and this is an aspect of behavior which is not given much attention in this book. The shortcoming may be serious if, as I believe, the wife's decisions about labor force activity and childbearing are in part "causes" and in part "effects" of the various

outcomes of husbands which the authors examine. A fuller treatment of the wife's behavior would challenge, to cite just one example, the authors' estimate of the effect of fertility on the husband's income and occupational status. (See section 8.5.) Although the authors indicate one reason why causation runs from fertility to income (and occupation), a common alternative view in economic models is that the husband's income bears a net negative relation to the wife's labor force activity and that wives who work less are likely to have more children.[8] Incidentally, the wife's labor force and fertility decisions are usually considered to be "simultaneous."

Another useful distinction in applied econometric work is that between variables which represent potential policy instruments and those that are not amenable to policy manipulation. This perspective leads to a meaningful basis for asking the question of whether the effect (coefficient) of the variable is "large" or "small," "important" or "unimportant." As argued elsewhere (Cain and Watts 1970), regression coefficients provide a necessary ingredient to answering these questions.[9] Note that this claim for the policy relevance of the regression coefficients is separate from the claim that they have greater stability than correlation coefficients or beta-weights. (See Tukey 1954.) The book's focus on path coefficients, simple and partial correlations, and "contribution" to the multiple R^2 provides statistics which are inherently ambiguous for assessing and ranking the importance of variables in multiple regression models.

Many examples could be cited to illustrate this ambiguity. In Chapter 7 the authors reveal that the effect of the wife's education on her husband's occupational status is positive and statistically significant. They say, however, that "wives apparently contribute little directly to their

husband's career--at least by way of their socioeconomic characteristics and personality traits" (p. 179). Setting to one side the questionable causal nature of the wife's education with respect to husband's occupation, we have a right to ask: What is a "little" effect here? The authors also say that the wife's education relates "slightly more importantly" to husband's occupational status than does his father's occupational status. What justifies this claim? Does marrying a wife with a college degree instead of a high school degree have a large effect on husband's occupation compared to his having a father with a professional, as opposed to a craftsman, occupation? Apparently the authors' answers depend on a ranking of the variables according to the sizes of their path coefficients. But they do not explain how the sizes of path coefficients, which embody a mixture of the quantitative relations (regression coefficients) and the sample variances, are useful measures of a variable's importance.

In another example, school inputs are downgraded as a determinant of educational aspirations on the basis of an analysis devoted to partitioning variances into "within" and "between" school components (pp. 191-198). But this lengthy and involved exercise says nothing about the policy relevant issue of the quantitative effects on educational outcomes of changes in school inputs. It is not clear whether the authors address this question, given their fixation on accounting for explained variation (i.e., for the R^2). I am surprised to see this fixation, since one of the authors has written an excellent article on the limitations of variance partitioning (Duncan 1971). My guess is that if the purposes for which the model was estimated were made explicit, many of these issues would be clarified.

This section of the review has been monopolized by an economist's perspective. Let me close with one example of a disputed interpretation

that more directly involves sociologists and cultural anthropologists. In analyzing black-white differences in income in Chapter 4, the authors, in effect, attribute this gap to two sources which are defined by the equations determining income: one source is the different values of the variables for the two racial groups, and the other source is the differences in parameters (effects) of those variables. When they equate the values of the variables, specifically the values for the person's father's socioeconomic level, they find that only 22 percent of the income gap is accounted for. They assert that this estimate "should give pause to those who cite the 'vicious circle of poverty' or the 'culture' of poverty as the fundamental cause of the black's lower income" (p. 60). One can agree with the authors' skepticism about these explanations and yet disagree that the values of the variables rather than the values of the parameters are what the exponents of the "culture of poverty" have in mind.

5. CONCLUSIONS

This outstanding book represents a milestone in the grand research tradition of the development of quantitative models of the processes of achievement and social stratification. As promised it provides "systematization and synthesis," "interpretation and generalization," and in so doing it transcends descriptive fact presentation or, at the other extreme, arm-chair theorizing. There is hardly a section that does not stimulate and challenge the reader with some interesting empirical relation; some skillful use of statistical technique; some enlightening comment on the methodology of measurement, model construction, and estimation in the social sciences; some insightful comment leading to, in their words, "a clearer vision of the entire process of social stratification."

The criticisms and questions raised in this review are intended to suggest ways in which this type of research could reach a wider audience and contribute more effectively to the application of social science to coping with social problems.

FOOTNOTES

[1]A major, earlier book (Morgan and others, 1952), not referred to by the authors, offers a number of similar models for explaining income and educational attainment, among other outcome variables, by adults in 1959. (See, in particular, Chapter 24 and Appendix E.) Two advantages of this study over the book under review are: (1) the quantitative effects of father's education, father's occupation, number of siblings, other background factors, and several exogenous personal variables of the adult are revealed in terms of the regression coefficients of the "natural units" of the variables; (2) the functional forms of the variables are not restricted to be linear and additive.

[2]The numerical example above is hypothetical. Actually, in the first set of empirical results the direct effect of the father's education on the son's occupation is only about one-sixth this size, and is, in fact, labeled a zero effect in the diagram (p. 39) and text discussion. Clearly, larger changes in occupational scores would reveal sharper and more meaningful contrasts in the occupations, and it is not my point to dwell on the difficulty of evaluating "small" changes.

[3]The prestige score, Y', was actually based on scores derived from a 1964 survey of the National Opinion Research Center, rather than from the original 1947 survey by NORC. However, this is irrelevant to the points made above.

[4]Clearly, x_1 will be positively correlated with both x_3 and x_6, and x_3 will be positively correlated with x_6. These results guarantee that $q_{63} > p_{63}$ except in the implausible case--especially implausible as an empirical matter--in which x_1 affects x_6 only through its effect on the empirical measure of educational aspirations (x_3).

[5]The source for obtaining assumed values of the path coefficients is a set of correlations based on a model in which SA is determined by the conventional variables, IGA is determined by A and SA, and CW is determined by WO and SA. (See p. 132.) Different assumptions about the reliability of the correlations involving A and WO provide several values of the path coefficients.

[6]These restrictive assumptions are not explicit in the authors' model using the unobservable M, but they are explicit in the model which uses the observable variables. The authors' model and the reformulated model are the same except for the definitional substitution, $X_M = (X_4/p_{4M}) + e$. The term X_4 is observable while the other terms are not.

[7]A useful discussion of this point and, indeed, of several issues that appear in this review is provided by Goldberger (1973).

[8]If the hypothesis (not necessarily economic) that higher income permits the parents to afford more children is true, then this would be another source of a causal relation running from income to fertility. A test of this hypothesis would, as the authors point out, require that current income and current occupation be good proxies for the theoretical variables, the "permanent" income and occupation which the parents expect to have (or attain).

[9]The standard errors of regression coefficients are also necessary to assess the statistical significance of the effect and to permit confidence limits to be established. The authors seldom report the standard errors of the path coefficients and sometimes do not report whether the coefficients (or correlations) are statistically significant.

REFERENCES

1. Blau, P. M. and Duncan, O. D., The American Occupational Structure, New York: Wiley, 1967.

2. Cain, G. G. and Watts, H. W., "Problems in Making Policy Inferences from the Coleman Report," American Sociological Review, April, 1970, pp. 228-42.

3. Duncan, O. D., "A Socioeconomic Index for All Occupations," in A. J. Reiss and others, Occupations and Social Status, New York: Free Press, 1961, pp. 109-138.

4. Duncan, O. D., "Partials, Partitions, and Paths," in E. Borgatta, ed., Sociological Methodology: 1970, San Francisco: Jossey-Boss, 1970, pp. 38-47.

5. Goldberger, A. G., "Structural Equations Models: An Overview," in Goldberger, A. G. and Duncan, O. D., eds., Structural Equation Models in the Social Sciences, New York and London: Seminar Press, 1973, pp. 1-18.

6. Morgan, J. and others, Income and Welfare in the United States, New York: McGraw-Hill, 1962.

7. Tukey, J. W., "Causation, Regression, and Path Analyses," in O. Kempthorne and others, eds., Statistics and Mathematics in Biology, Ames: Iowa State University Press, 1954.